

Обучение машинного перевода без параллельных текстов*

Скиднов Е. А. Бахтеев¹ О. Ю. Стрижов² В. В.

¹Московский физико-технический институт

²Вычислительный центр им. А. А. Дородницына ФИЦ ИУ РАН

Данная задача посвящена исследованию алгоритма обучения машинного перевода без параллельных предложений. Использование параллельных текстов для задачи машинного перевода требует слишком большой базы предложений всех переводимых языков, что является ресурсоемкой задачей для некоторых пар непохожих языков. Особенностью исследуемого алгоритма является то, что для перевода используется кодирование и декодирование текста во внутреннем представлении. Данный алгоритм использует 2 модели нейронной сети Seq2Seq для перевода с одного языка на другой и обратно. Цель данного исследования заключается в том, чтобы сделать вектора скрытых пространств этих двух моделей как можно более похожими. Для демонстрации работоспособности метода будет использован вычислительный эксперимент машинного перевода между двумя похожими языками: русским и украинским.

Ключевые слова: *машинный перевод, нейросеть, Seq2Seq.*

*Работа выполнена при финансовой поддержке РФФИ, проект № 00-00-00000. Научный руководитель: Стрижов В. В. Авторы: А. В. Грабовой, О. Ю. Бахтеев, В. В. Стрижов, Eric Gaussier, координатор Малиновский Г. С. Консультант: Бахтеев О. Ю.